JOURNAL OF ECONOMIC THEORY 23, 150–159 (1980)

Stability of Decision Systems under Majority Rule

ARIEL RUBINSTEIN

Department of Economics, The Hebrew University, Jerusalem, Israel

Received August 10, 1977; revised November 8, 1979

1. INTRODUCTION

As the well-known "Paradox of Voting" indicates, a decision system under majority rule does not necessarily have an equilibrium; that is, it might be that for every alternative there is a majority preferring another. Many writers, particularly political scientists, thus inferred the weakness of majority-rule systems (see, for example, Brams (1976), Riker and Ordeshook (1973)). A social system with no equilibrium might be thrust into continuous cyclic change. Moreover, the individual members who believe in democracy would be permanently dissatisfied, since they are always convinced (and correctly so), that there is a majority in favor of a social reality different from the existing one.

Many attempts have been made to find sufficient conditions for the existence of equilibrium in systems under majority rule, for example, Arrow (1963), Davis and Hinich (1972), Dommet and Farquharson (1961), Nakamura (1975), and Plott (1967). The prevalent impression arising from reading these works is that only strong conditions on the decision system lead to the existence of equilibrium. (This impression is expressed topologically in Rubinstein (1979).)

The starting point of all these papers is the definition of equilibrium and the identification of social stability with the existence of equilibrium. The existence of a majority preferring b to a deprives a of stability. The individuals are assumed not to take into account what happens after the system switches to b.

In this paper I will attempt to see if the general picture is less pessimistic when we assume that the individuals are aware of possible future developments. The behavior pattern examined is that determined by the following reasoning: "True, I prefer b to a, but if b is adopted, then a situation arises where the majority prefers c. Since c is worse than a from my point of view, I will not take any chances and will not vote for b in place of a." A social possibility will be considered to be stable if no majority exists for change when all the individuals adopt this more farseeing behavior

0022-0531/80/050150–10$02.00/0

pattern. Further analyses are needed to investigate behavior patterns which are intermediate to the unstrategic behavior which gives rise to the "Paradox of Voting" and that assumed here (proper solution concepts may be those of Vickrey (1953)).

There are two main parts to this paper following the description of the model and the definitions of the stability set, J (Section 2). In the first part two nonemptiness theorems are proved (Section 3). In the second part, the concept of dynamic system as defined by Maschler and Peleg (1976) is used to examine the stability of J in the dynamic system corresponding to the behavior pattern described (Section 4).

2. The Model

A social decision system will be formulized by an ordered triplet

$$\langle A, \{\lesssim_i\}_{i=1}^{n}, W\rangle,$$

where

(1) A is the set of social states.

(2) $N = \{1, \ldots, n\}$ is the set of individuals and $\lesssim_i$ is the preference relation of individual i. The relation $\lesssim_i$ is assumed to be a reflexive, connected, and transitive relation.

(3) The institutional method by which the society makes decisions is denoted by $W \subseteq 2^N \times 2^N$ (2^N is the power set of N). We require that

(i) $W \neq \phi$.

(ii) $(S_1, S_2) \in W \Rightarrow S_1 \cap S_2 = \phi$.

(iii) If $(S_1, S_2) \in W$, $S_1 \cup S_2 \subseteq T_1 \cup T_2$, $S_1 \subseteq T_1$, and $T_1 \cap T_2 = \phi$, then $(T_1, T_2) \in W$.

(iv) If $(S_1, S_2) \in W$ then $(N - S_1 - S_2, \phi) \notin W$.

The interpretation of W is that "$(S_1, S_2) \in W$" when, for any two choices a and b, if the individuals in S_1 vote for b and the individuals in S_2 abstain, then the society will adopt b whatever $N - (S_1 \cup S_2)$ do. Here are some examples of possible sets W within our framework:

(a) Dictatorship. There exists i_0 such that $W = \{(S_1, S_2) \mid i_0 \in S_1\}$.

(b) An absolute majority is required to bring change.

$$W = \left\{(S_1, S_2) \mid |S_1| > \frac{n}{2}\right\}.$$

(c) A relative majority is required to bring change.

$$W = \{(S_1, S_2) \mid |S_1| > |N - S_1 - S_2|\}.$$

(d) Same as (c), but a chairman i_0 has the casting vote.

$$W = \{(S_1, S_2) \mid |S_1| > |N - S_1 - S_2| \text{ or } |S_1| = |N - S_1 - S_2| \text{ and } i_0 \in S_1\}.$$

DEFINITIONS. (1) For all $S \subseteq N$, $<_S$ is defined by

$$a <_S b \quad \text{if} \quad (\{i \in S \mid a <_i b\}, \{i \in S \mid a \sim_i b\}) \in W.$$

($<_S$ describes the decisions the coalition S can enforce whenever all the members of S behave honestly)

(2) $<$ is defined by

$$a < b \quad \text{if there exists } S \subseteq N \text{ such that } a <_S b.$$

(3) The core, C, is defined by

$$C = \{a \in A \mid \text{no } b \in A \text{ satisfies } a < b\}.$$

An element of C will be called an equilibrium of the system.

(4) $a \lll_S b$ if $a <_S b$ and there does not exist $i \in S$ and $b < c$ such that $c <_i a$.

(5) $a \lll b$ if there exists S such that $a \lll_S b$.

(6) The stability set, J, is defined by

$$J = \{a \in A \mid \text{no } b \in A \text{ satisfies } a \lll b\}.$$

Thus the stability set is the set of all the social alternatives for which no coalition, adopting the above behavior pattern, can by honest voting enforce another alternative.

The model is also suitable to describe a decision-making institution with two identical houses, a lower house and an upper house. The lower house decides whether to replace a by b. If it accepts b, the upper house may modify b and replace it by c which is the final decision. Clearly members of the lower house will think twice before voting for b. They might first check if the upper house will contain a majority in favor of modifying b to c, which might be worse for them than a. The stability set is the set of equilibria whenever no enforcible contracts are allowed between members of the two houses and the members of the lower house adopt the above very careful behavior pattern.

EXAMPLES. (1) Let

$$A = \{a, b, c\},$$
$$W = \{(S^+, S^=) \mid |N - S^+ - S^=| < |S^+|\},$$
$$N = \{1, 2, 3\},$$

and suppose

$$c <_1 b <_1 a$$
$$a <_2 c <_2 b$$
$$b <_3 a \sim_3 c.$$

Then $b <_{\{1,3\}} a$ $c <_{\{2,3\}} b$ and $c \not< a$ and $a \not< c$. Thus $C = \{a\}$ but $J = \{a, c\}$.

(2) The stability set may contain points which are not Pareto-optimal. Denote $P = \{a \in A \mid \text{there is no } b \in A,\ a <_i b \text{ for all } i \in N\}$.

Let

$$A = \{a, b, c, d\},$$
$$W = \{(S^+, S^=) \mid |N - S^+ - S^=| < |S^+|\},$$
$$N = \{1, 2, 3\},$$

and suppose

$$a <_1 b <_1 c <_1 d$$
$$c <_2 a <_2 b <_2 d$$
$$d <_3 a <_3 b <_3 c.$$

Then $a \in J$ but $a \notin P$.

This example demonstrates a weakness of our solution concept. It does not support the following line of reasoning which would remove a from the stability set. "If I support the replacement of a by b, then I might arrive at c, which is worse for me than a. However, c may be reached from a as well, so I may as well support b."

3. NONEMPTINESS THEOREMS

THEOREM 3.1. *Let $\langle A, \{\lesssim_i\}_{i=1}^n, W\rangle$ be a social decision system, where A is finite and for all i, $\lesssim_i$ is a linear order (also antisymmetric). Then the stability set is nonempty.*

Proof. For all $a \in A$, define $F(a) = |\{b \mid a < b\}|$. Let a_0 satisfy $F(a_0) = \min_{a \in A} F(a)$. Let b and S satisfy $a_0 <_S b$. The relation $<$ is asymmetric due to the conditions on W. Therefore there exists $c \in A$ ($c \neq b$, $c \neq a_0$) such that $b < c$ and $a_0 \not< c$. Thus there exists $i \in S$ such that $c <_i a_0$. Therefore $a_0 \in J$ and $J \neq \phi$.

Clearly if A is infinite, J may be empty (for example, in the case where A is the set of the natural numbers and the individuals' preferences is the standard order). I will now prove a nonemptiness theorem for social decision systems, where for all i, $\lesssim_i$ is a well ordering (i.e., for any B, a subset of A, there exists $\bar{b} \in B$ such that $b \lesssim_i \bar{b}$ for all $b \in B$).

THEOREM 3.2. *Let $\langle A, \{\lesssim_i\}_{i=1}^n, W\rangle$ be a system satisfying*

(i) *A is an infinite set.*

(ii) *For all $1 \leqslant i \leqslant n$, $\lesssim_i$ is a linear well ordering.*

Then the stability set is nonempty.

LEMMA. *Let R and S be two linear well orderings on an infinite set X. Then there is an infinite set Y, $Y \subseteq X$, such that R and S are identical on Y.*

Proof. Define a graph[1] $\langle X, G\rangle$:

$$(a, b) \in G \qquad \text{iff} \qquad a\,R\,b \Leftrightarrow a\,S\,b.$$

Assume that there is no full[2] infinite subgraph of $\langle X, G\rangle$. Then, there is an infinite empty subgraph of $\langle X, G\rangle$ (see Chang and Keisler (1973)); denote this subgraph $\langle G, G|_C\rangle$. Every subset of C has a maximal member according to the relation R which is therefore the minimal member according to S. Thus C is well ordered from above and below by S and is therefore a finite set.

COROLLARY. *P, the set of Pareto-optimal alternatives, is finite and nonempty.*

Proof. Assume P is infinite. From the Lemma there is an infinite subset of P on which all the individuals' preferences are identical, in contradiction to the definition of P.

Proof of the theorem. From 3.1 the stability set of $\langle P, \{\lesssim_i|_P\}_{i=1}^n, W\rangle$ is not empty. Let p be a member of the stability set. It will be proven that p is also a member of the stability set of the original social decision system. Assume $p \ll_S q$. Without loss of generality $q \in P$ (otherwise choose an $<_1$-maximal of $\{t \mid q <_i t\ \forall i\}$). From the choice of p follows the existence of $t \in P$ and $i \in S$ such that $q < t$ and $t <_i p$.

[1] A graph is a pair (C, G), where $G \subseteq \{K \subseteq C \mid |K| = 2\}$.

[2] $\langle C, G\rangle$ is a full graph if $G = \{K \subseteq C \mid |K| = 2\}$. $\langle C, G\rangle$ is an empty graph if $G = \phi$.

4. Stability Theorem

Let us begin with definitions and notation which are taken from Maschler and Peleg (1976).

A (set-valued) dynamic system is an ordered pair $\langle X, \varphi \rangle$, where X is metric space and $\varphi: X \to 2^X$ satisfies $\varphi(x) \neq \phi$ for all $x \in X$.

A φ-sequence (starting at x^0) is a sequence (x^t) such that $x^0 \in X$ and $x^{t+1} \in \varphi(x^t)$ for $t = 0, 1, 2,...$.

A point $x \in X$ is called an endpoint of φ if $\varphi(x) = \{x\}$. The set of endpoints of φ, will be denoted by $E(\varphi)$.

A nonempty subset of X, Q, is called stable with respect to φ if for every neighborhood U of Q there exists a neighborhood V of Q such that if (x^t) is a φ-sequence starting in V then $x^t \in U$ for all $t \geqslant 0$. A point $x \in X$ is called stable if $\{x\}$ stable set.

Let $g(x) = (G_1(x),..., G_m(x))$ be a vector of m real-valued functions on X. A point $a \in R^m$ is called Pareto-minimal with respect to g if there exists $x \in X$ such that $g(x) = a$ and if whenever $y \in X$ and $G_i(y) \leqslant a_i$ for $i = 1,..., m$, then $g(y) = a$.

Let a be Pareto-minimal with respect to g. The set $Nu(g, a) = \{x \in X \mid g(x) = a\}$ is called nucleolus of g with respect to a.

g is called φ-monotone if for all $x \in X$, $y \in \varphi(x) \Rightarrow G_i(x) \geqslant G_i(y)$ $i = 1,..., m$. g is called strictly φ-monotone if it is φ-monotone and for all $x \in X$ and $y \in \varphi(x)$, $y \neq x \Rightarrow G_k(x) > G_k(y)$ for some $1 \leqslant k \leqslant m$.

In the following, the discussion is restricted to social systems satisfying:

(A.1) A is a nonempty, convex, and compact subset of E^k.

(A.2) For any i, $\lesssim_i$ is a strongly convex[3] continuous relation.

(A.3) For all (R_1, R_2), $(S_1, S_2) \in W$ $(R_1 \cup R_2) \cap S_1 \neq \phi$.

(For example, this condition is satisfied by examples a, b, d in Section 2.) The above assumptions have often appeared in the literature, and have been followed by "pessimistic" conclusions (see, for example, Davis and Hinch (1972) and Plott (1967)).

Theorem 4.1. *Let $\langle A, \{\lesssim_i\}_{i=1}^n, W \rangle$ be a social system satisfying* (A.1), (A.2), *and* (A.3). *Let $\langle A, \varphi \rangle$ be a dynamic system satisfying*

$$\varphi(x) = \{x\} \qquad \text{if there is no } y \in A,\ x \ll y$$
$$= \{y \mid x \ll y\} \qquad \text{otherwise.}$$

Then $J \neq \phi$, and there exists $\bar{J} \subseteq J$ such that $\bar{J}$ is stable relative to φ.

[3] If $a \lesssim_i b$ and $a \lesssim_i c$ and $b \neq c$ then for any $0 < \lambda < 1$ $a <_i \lambda b + (1 - \lambda)c$.

Proof. Assumptions (A.1), (A.2), and (A.3) guarantee the existence of continuous, single-peaked strict quasi-convex loss function $\{d_i\}_{i=1}^n$ corresponding to $\{\lesssim_i\}_{i=1}^n$ $(x \lesssim_i y \Leftrightarrow d_i(x) \geqslant d_i(y)$, there exists a unique $v_i \in E^k$ such that $d_i(v_i) = 0$ and for all $a \neq b$ and $0 < \lambda < 1$, $d_i(a) \leqslant d_i(b) \Rightarrow d_i(\lambda b + (1-\lambda)a) < d_i(b))$. The loss functions $\{d_i\}_{i=1}^n$ will be fixed for the rest of the proof.

Define a vector function $(g_R)_{\phi \neq R \subseteq N}$ which will serve as a Liaponov function for the dynamic system as follows: for all $\phi \neq R \subseteq N$, $g_R : A \to E^1$ and

$$g_R(a) = \sup_{a \prec b} \sum_{i \in R} d_i(b) \qquad \text{if there exists } a \prec b$$

$$= \sum_{i \in R} d_i(a) \qquad \text{otherwise.}$$

d_i are bounded (continuous functions on a compact set) and thus $0 \leqslant g_R(a) < \infty$. Denote $g_{\{i\}}$ by g_i.

LEMMA A. $\sum_{i \in R} d_i(a) \leqslant g_R(a)$.

Proof. If $a \in C$ $g_R(a) = \sum_{i \in R} d_i(a)$. Otherwise, let $a \prec_T b$. From the strong quasi-concavity of the individuals' preferences, for any $0 < \varepsilon < 1$ and for any $i \in T$ $a \prec_i \varepsilon a + (1-\varepsilon)b$ and from assumption (iii) on W $a \prec_T \varepsilon a + (1-\varepsilon)b$. Therefore $\sum_{i \in R} d_i(\varepsilon a + (1-\varepsilon)b) \leqslant g_R(a)$ and together with the continuity of the loss functions this implies the Lemma.

LEMMA B. *g is φ-monotone.*

Proof. Let $b \in \varphi(a)$ and let us assume $a \preccurlyeq_T b$. If there is no c such that $b \prec c$, then $g_R(a) \geqslant \sum_{i \in R} d_i(b) = g_R(b)$. If there exists c such that $b \prec c$, then for every $i \in T$, $a \lesssim_i c$, and from the strong convexity it follows that $a \prec_i \varepsilon a + (1-\varepsilon)c$ for all $0 < \varepsilon < 1$ and for all $i \in T$. Then assumption (iii) on W assures that $a \prec_T \varepsilon a + (1-\varepsilon)c$ and $\sum_{i \in R} d_i(\varepsilon a + (1-\varepsilon)c) \leqslant g_R(a)$. d_i $(i = 1,..., n)$ are continuous, so when $\varepsilon \to 0$ $\sum_{i \in R} d_i(c) \leqslant g_R(a)$. This holds for all $b \prec c$. Thus $g_R(b) \leqslant g_R(a)$.

LEMMA C. *g is strictly φ-monotone.*

Proof. Let $b \in \varphi(a)$ and suppose $a \lll_T b$. From the assumptions on W there exists $i \in T$ such that $a \prec_i b$, and if $b \in C$

$$g_i(b) = d_i(b) < d_i(a) \leqslant g_i(a).$$

If $b \notin C$, let $b \prec_S c$. Let i_0 be a member of S, which also satisfies $a \prec_{i_0} b$ (the existence of such an i_0 follows from A.3). Thus,

$$\sum_{i \in T} d_i(c) = \sum_{i \in T - \{i_0\}} d_i(c) + d_{i_0}(c) \leqslant \sum_{i \in T - \{i_0\}} d_i(a) + d_{i_0}(b)$$
$$\leqslant \sum_{i \in T - \{i_0\}} d_i(a) + d_{i_0}(a) - \varepsilon = \sum_{i \in T} d_i(a) - \varepsilon$$
$$\leqslant g_T(a) - \varepsilon,$$

where $\varepsilon = \min_{i \in T^+} d_i(a) - d_i(b) > 0$.
Since c is arbitrary and ε is independent of c,

$$g_T(b) < g_T(b) + \varepsilon \leqslant g_T(a).$$

LEMMA D. *For any $\varnothing \neq R \subseteq N$, g_R is lower semicontinuous.*[4]

Proof. Let $a_n \to a$. If $a \in C$, then for every $i \in N$ $g_i(a) = d_i(a) \leftarrow d_i(a_n) \leqslant g_i(a_n)$, hence $g_i(a) \leqslant \underline{\lim}\, g_i(a_n)$.

Suppose for some $b \in A$ $a \prec_S b$. Let $0 < \varepsilon < 1$ and $j \in S$. From the strong convexity it follows that $a \prec_j \varepsilon a + (1-\varepsilon)b$, and by the continuity of $\precsim_j$, there exists n_0 such that for every $n_0 < n$ $a_n \prec_j \varepsilon a + (1-\varepsilon)b$. So far large anough n, $a_n \prec \varepsilon a + (1-\varepsilon)b$. $g_R(a_n) \geqslant \sum_{i \in R} d_i(\varepsilon a + (1-\varepsilon)b)$; consequently $\underline{\lim}\, g_R(a_n) \geqslant \sum_{i \in R} d_i(\varepsilon a + (1-\varepsilon)b)$. Letting $\varepsilon \to 0$, and by the continuity of the d_i, we obtain $\underline{\lim}\, g_R(a_n) \geqslant \sum_{i \in R} d_i(b)$ and thus $\underline{\lim}\, g_R(a_n) \geqslant g_R(a)$.

LEMMA E. *For any $\phi \neq R \subseteq N$ g_R is upper semicontinuous.*

Proof. Assume (a_n) is a sequence such that $a_n \to a$ and $\lim g_R(a_n) > g_R(a) + \varepsilon$. Clearly for almost all n, $a_n \notin C$. Thus w.l.o.g. assume that for all n $a_n \prec_T b_n$, $\sum_{i \in R} d_i(b_n) > g_R(a) + \varepsilon$ and $b_n \to b$. By continuity of (d_i), $\sum_{i \in R} d_i(b) \geqslant g_R(a) + \varepsilon$. From Lemma A, $b \neq a$ and from the continuity of $(\precsim_i)$, for all $i \in T$, $a \precsim_i b$. By the strong quasi-convexity of $(\precsim_i)$, for all $0 < t < 1$ and $i \in T$. $a \prec_i tb + (1-t)a$ and as $t \to 1$ $\sum_{i \in R} d_i(b) \leftarrow \sum_{i \in R} d_i(tb + (1-t)a) \leqslant g_R(a)$, leading to a contradiction.

Proof of the theorem. Clearly if g is strictly φ-monotone every $Nu(g, a)$ is a subset of $E(\varphi) = J$. X is a compact metric space, g is φ-monotone, and (g_R) are continuous, so $Nu(g, a)$ is stable and closed (Maschler and Peleg (1976)).

[4] A real-valued function f is lower (upper) semicontinuous if for any a and $a_n \to a$, $\underline{\lim}\, f(a_n) \geqslant f(a)$ ($\overline{\lim}\, f(a_n) \leqslant f(a)$).

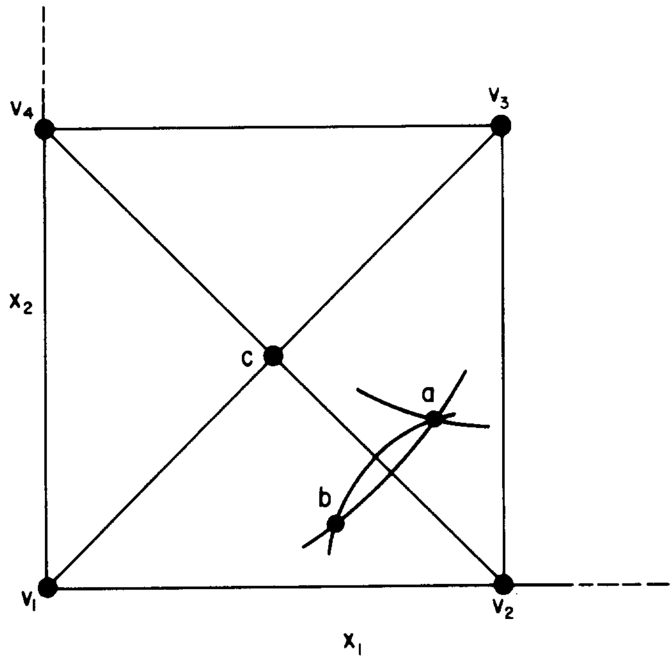

FIGURE 1

An Example for Calculating the Nucleolus.

Let

$$N = \{1, 2, 3, 4\},$$
$$A = \text{The unit square},$$
$$W = \{(S^+, S^=) \mid |N - S^+ - S^=| < |S^+|\},$$

and for all $i \in N$

$$d_i(x) = \|x - v_i\| \text{ (see Figure 1).}$$

Let us calculate $g_1(x)$. It is readily ascertained that the worst possibility for 1 in either of the states a and b is a; from this we have

$$\begin{aligned} g_1(\langle x_1, x_2 \rangle) &= (x_1{}^2 + x_2{}^2)^{1/2} && \text{if} \quad x_1 + x_2 \geqslant 1 \\ &= ((1 - x_1)^2 + (1 - x_2)^2)^{1/2} && \text{if} \quad x_1 + x_2 \leqslant 1. \end{aligned}$$

The only Pareto-minimal point obtained is

$$a = \langle \underbrace{\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}}_{|s|=1}, \underbrace{1,\dots, 1}_{|s|=2}, \underbrace{\tfrac{3}{2},\dots, \tfrac{3}{2}}_{|s|=3N}, 2\rangle,$$

and the nucleolus is $\{i\}$.

Acknowledgments

My thanks are due to Professor M. E. Yaari of the Department of Economics and to Professor B. Peleg of the Institute of Mathematics at the Hebrew University of Jerusalem, for their comments. Of course, any errors are my sole responsibility.

References

K. J. Arrow (1963), "Social Choice and Individual Values," 2nd ed., Yale Univ. Press, New Haven, Conn.

S. J. Brams (1976), "Paradoxes in Politics," Free Press, New York.

C. L. Chang and H. J. Keisler (1973), "Model Theory," North-Holland, Amsterdam.

O. A. Davis and M. J. Hinich (1972), Social preference orderings and majority rule, *Econometrica* **40**, 147–157.

M. Dommet and R. Farquharson (1961), Stability in voting, *Econometrica* **29**, 33–43.

M. Maschler and B. Peleg (1976), Stable sets and stable points of set-valued dynamic systems with applications to game theory, *SIAM J. Control Optimization* **14**, 985–995.

K. Nakamura (1975), The core of a simple game with ordinal preferences, *Internat. J. Game Theory* **4**, 95–104.

C. R. Plott (1967), A notion of equilibrium and its possibility under majority rule, *Amer. Econ. Rev.* **57**, 787–806.

W. H. Riker and P. C. Ordeshook (1973), "An Introduction to Positive Political Theory," Prentice-Hall, Englewood Cliffs, N.J.

A. Rubinstein (1979), A note about the "nowhere denseness" of societies having an equilibrium under majority rule, *Econometrica* **47**, 511–514.

W. Vickrey (1953), Strong and weak solutions in the theory of games, unpublished.